

CIVIC DIGITAL FELLOWSHIP

Automating User Feedback Analysis for data.census.gov

Dylan Bragdon

Elijah Tai

Supervised by Zachary Whitman

Center for Enterprise Dissemination, Research and Methodology Division

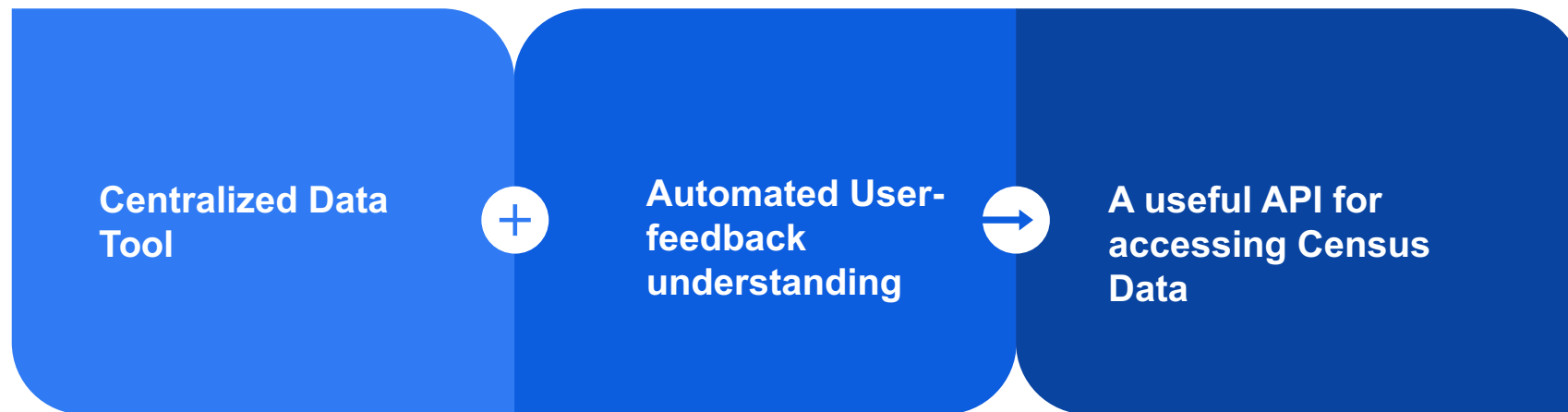
Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau.

Shape
your future
START HERE >

United States[®]
Census
2020

Problem and Significance

- data.census.gov: “The new platform for accessing data and digital content from the U.S. Census Bureau”
- New tool, many problems
- User-feedback for data.census.gov is currently hand-labeled
- Can we streamline a process for labeling and categorizing user feedback?



Shape
your future
START HERE >

United States[®]
Census
2020

Issues with Old Approach

- Much of the hand-labelled data is incorrect
- Hand labelling new data is time consuming
- Current labels are uninformative or highly correlated
- Current topics are too broad

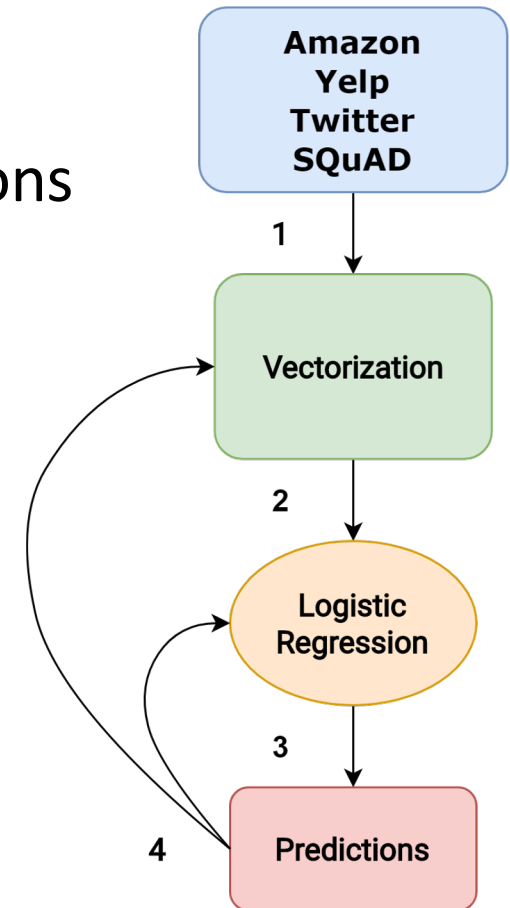
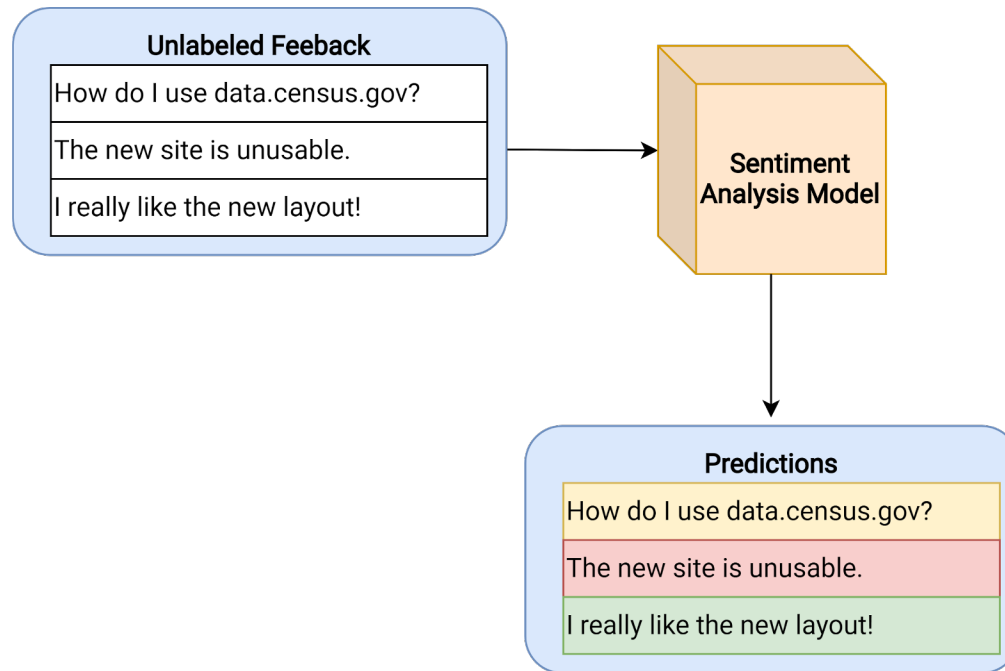
Project Objectives

- Create a dependable model for classifying the **sentiment** of user feedback
- Extract common points of frustration from users of data.census.gov

1	2	3	4
Sentiment Analysis Prerequisites Gather external data, determine informative labels.	Tune Hyperparameters Adjust model for optimal performance on training set.	Work with user feedback Applying sentiment analysis model on real data, extracting topics.	Future data Create a pipeline of feature extraction for future user feedback.

Sentiment Analysis

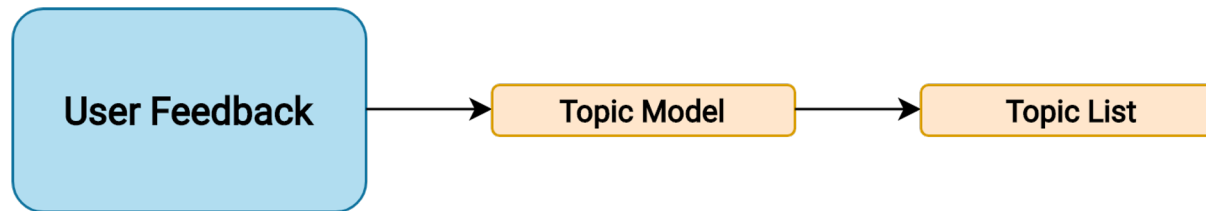
- Trained on external, multi-domain data
- Data largely composed of compliments, complaints, and questions
- Tuned to fit the needs of both users and Census employees



Topic Modeling

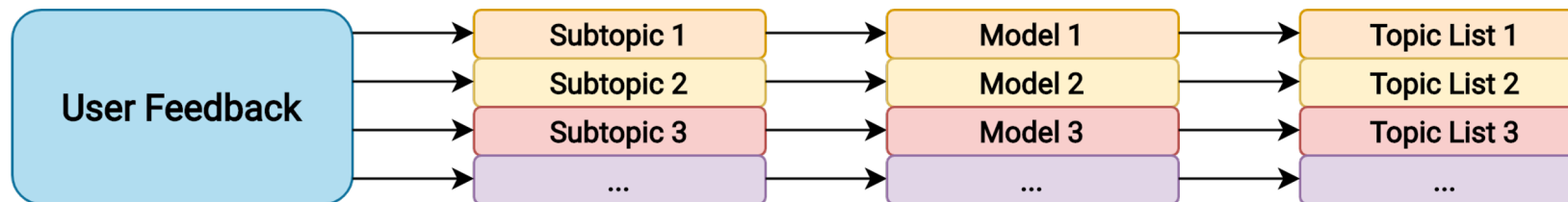
General Topics

→ Consider entire feedback history



Fine-grained Topics

→ Extract topics from pre-labelled subtopics



Deliverables

→ **Sentiment Analysis Model**

Fully Classified emails

→ **Topic Extraction Model**

Past feedback grouped by topic

→ **Insight into nature of user-feedback for data.census.gov**

Next steps: Improve models, expand label set,
create pipeline for new data